Multiple introductions of two invasive *Centaurea* taxa inferred from cpDNA haplotypes

Ruth A. Hufbauer[1]* and René Sforza[2]

[1]*Department of Bioagricultural Sciences and Pest Management and Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523-1177, USA,* [2]*European Biological Control Laboratory, USDA-ARS, Campus International de Baillarguet, C S90013, Montferrier-sur-Lez, 34988 St-Gely du Fesc, France*

*Correspondence: Ruth A. Hufbauer, Department of Bioagricultural Sciences and Pest Management, Colorado State University, 1177 Campus Mail, Fort Collins, CO 80523-1177, USA. Tel.: 970-491-6945; Fax: 970-491-3862; E-mail: ruth.hufbauer@colostate.edu

ABSTRACT

Knowing the origin of invasive taxa, whether multiple introductions have occurred, and levels of genetic variation relative to the native range, is vital to conducting rigorous tests of hypotheses to explain biological invasions. We explore phylogeographical relationships of two Eurasian knapweed taxa that are invasive in North America, *Centaurea diffusa* and *C. stoebe micranthos* (Asteraceae), using chloroplast DNA intron sequences. We also gathered data from *C. stoebe stoebe*, hybrids between *C. diffusa* and *C. stoebe stoebe* (*C. ×psammogena*), and three other species in the genus. We sequenced 213 individuals from Eurasia and North America, and found 22 haplotypes. *Centaurea diffusa* has lower haplotype diversity and allelic richness in the introduced range relative to the native range. Even with reduced variation, the data suggest at least two introductions of *C. diffusa*. There is a trend towards reduced variation in *C. stoebe micranthos*, but it is not significant. One of the haplotypes found in North American *C. stoebe micranthos* matches a haplotype from a taxon other than *C. stoebe micranthos* in Europe. This suggests introgression of the chloroplast between taxa, or possibly the invasion of another *Centaurea* taxon into North America. Additionally, *C. diffusa*, *C. stoebe micranthos*, and *C. stoebe stoebe* share several haplotypes, including their most common haplotype. This suggests ongoing hybridization between the species or incomplete segregation. These data can guide further exploration for the origins of these species, and point out locations within the introduced range with unique and diverse genetic makeup.

Keywords

Exotic plants, biological control, biological invasions, *Centaurea maculosa*, haplotype network, invasion, knapweed.

INTRODUCTION

Biological invasions are crucibles for rapid ecological and evolutionary change: they alter species distributions, ecosystem processes, and community structure, and are a key threat to native biodiversity (e.g. Parker *et al.*, 1999; Mack *et al.*, 2000; Pimentel *et al.*, 2000). Determining the factors that enable an introduced species to proliferate has emerged as a fundamental challenge to ecologists and evolutionary biologists worldwide (Thuiller *et al.*, 2005; Mitchell *et al.*, 2006; Pyšek & Richardson, 2006).

There are numerous ecological and evolutionary hypotheses to explain the invasiveness of non-native species (see for example, Shea & Chesson, 2002; Torchin & Mitchell, 2004; Richardson & Pyšek, 2006; Blank & Sforza, 2007). To test these hypotheses rigorously, we need to know what is present, where it was introduced from, and how genetically variable it is, all of which can be addressed initially with basic molecular population genetic surveys. While identifying what taxa have invaded may sound simple, often it is not (Jahodová *et al.*, 2007). Even within the native range, taxa may not be well-defined or described, and phenotypic differences between samples from the native and introduced range may make species identification difficult without the use of molecular markers. Molecular data can also aid in determining from where in the native range taxa were introduced, which information is required for rigorous tests of several hypotheses regarding invasion (e.g. the explicitly biogeographical enemy release and EICA hypotheses; Blossey & Nötzold, 1995; Torchin & Mitchell, 2004). Finally, molecular genetic data can be used to evaluate how much genetic variation is present in the introduced range, and how that differs from the native range, which is important in addressing ideas about how founder effects, hybridization, and outcrossing may inhibit or facilitate invasions (Ellstrand & Schierenbeck, 2000; Sakai *et al.*, 2001; Kolbe *et al.*, 2004; Facon *et al.*, 2006; Lavergne & Molofsky, 2007). In addition to enabling more rigorous tests of

ecological and evolutionary hypotheses, an important application of quantifying genetic variation is to focus work on biological control of invasive species. Evaluation of the efficacy of candidate agents for biological control should be done on the full range of genotypic diversity present within the introduced range (Gaskin *et al.*, 2005; Goolsby *et al.*, 2006), thus first that range of diversity must be delineated.

The genus *Centaurea* L. (Asteraceae) has about 300 species (Garcia-Jacas *et al.*, 2006), many of which have been introduced around the world and become invasive. In North America, 34 *Centaurea* species are reported to be introduced (http://plants.usda.gov/), 14 of which are defined as noxious weeds in one or more states. However, there is uncertainty regarding exactly which taxa are present in North America, and their proper scientific names. This uncertainty is compounded by the state of taxonomy of the group: sections within the genus are still being revised, and relationships within sections are not well resolved (Garcia-Jacas *et al.*, 2006).

Our research focuses on members of the *Centaurea* genus within the section Acrolophus–Phaelolepsis (Garcia-Jacas *et al.*, 2006). Furthermore, we focus on several members of the *Centaurea stoebe* species group (*C. stoebe* s.l.), a group of about 33 named taxa (Ochsmann, 2000). Two taxa within *C. stoebe* s.l., diffuse and spotted knapweed, have become noxious weeds in North America and are the main subjects of this research. Both of these taxa are comprised of diploid ($2x = 2n = 18$) and tetraploid ($2x = 4n = 36$) cytotypes. The cytotypes of diffuse knapweed fall under the single scientific name *Centaurea diffusa* Lam. The nomenclature for spotted knapweed is more complex. The two cytotypes, which are morphologically indistinguishable, together come under *C. stoebe* L, a name that takes precedence over the commonly used *C. maculosa* (Ochsmann, 2000). The monocarpic diploid is designated *C. stoebe* ssp. *stoebe* L., and the polycarpic tetraploid is designated C. *stoebe* ssp. *micranthos* (Gugler) Hayek (for which *C. biebersteinii* DC. is a synonym). Diploid *C. stoebe stoebe* are native to western Europe, and tetraploid *C. stoebe micranthos* are native to eastern Europe and Asia. The spotted knapweed plants in North America that have been surveyed are tetraploids (Müller, 1989; Ochsmann, 2000). Supporting these observations, most plants within North America are polycarpic (RAH, pers. obs.), and have up to four alleles at microsatellite loci (Marrs *et al.*, 2006). Thus, we call our North American samples *C. stoebe micranthos*. Diploid hybrids between diploid *C. diffusa* and *C. stoebe stoebe* were first described in the native range in 1909 (Gáyer, 1909), and given the name *C.* ×*psammogena* Gáyer. Plants with morphological characteristics of hybrids are also found in North America.

Centaurea diffusa and *C. stoebe micranthos* were first recorded in the USA in Washington State in the early 1900s. They were introduced from Eurasia as contaminants of alfalfa seed or in ballast (Maddox, 1979). Since then they have spread rapidly across western North America (Roché & Roché, 1991; Sheley *et al.*, 1999) and western Canada (Watson & Renney, 1974). Both species have become models for research on the ecological causes and consequences of invasions (e.g. Pearson *et al.*, 2000; Story *et al.*, 2000; Seastedt *et al.*, 2003; Callaway *et al.*, 2004; Suding

et al., 2004). Despite this attention, little is known about the genealogical relationships between individuals from the native and introduced ranges. Even more importantly, with the complexity of the genus, the taxonomic identity of the individuals in North America remains uncertain. Positive identification is complicated by the diversity of taxa within the *Centaurea* genus in the native range, many of which can be difficult to distinguish morphologically (Ochsmann, 2000).

To move towards rigorous tests of hypotheses to explain and manage biological invasions, we are exploring the population genetics of the invasion of *C. diffusa* and *C. stoebe micranthos* into North America. We gathered sequence data from chloroplast introns to evaluate the region's utility and to explore: (1) whether the two species were introduced one or more times into North America, (2) possible origins in the native range of invasive individuals, and (3) levels of genetic variation within the native and introduced ranges.

METHODS

Plant collections and DNA extractions

Either fresh leaf tissue or seed were collected from 1 to 30 individual plants from 30 locations in the native European range and 20 locations in the introduced North American range (Fig. 1, Table 1). Eurasian samples described here were identified following Turkish and Ukrainian floras, and species assignments were verified by botanists from herbaria in Ukraine (Kiev and Yalta), and France (Paris and Montpellier). Leaf tissue was stored in silica desiccant until DNA could be extracted. Seeds were stored in paper coin envelopes, and later germinated in a greenhouse. At approximately the four-leaf stage, one leaf was harvested for DNA extraction. Genomic DNA was extracted using the Dneasy Plant Mini Kit from Qiagen® (Valencia, CA, USA). This study included a total of 213 samples (Table 1). Initially, all plants sampled were thought to be *C. diffusa*, *C. stoebe stoebe*, *C. stoebe micranthos*, *C. maculosa albida* Lecoq & Lemotte (distinct populations that may be synonymous with *C. stoebe stoebe*), *C. vallesiaca* D.C. Jord. (a closely related species within the *C. stoebe* s.l. species group from its native Swiss Valais), and *C.* ×*psammogena*. Subsequent identification of herbarium specimens, however, indicated that samples from two of the locations in the native range that were thought to be *C. stoebe micranthos* instead represented two additional species. Samples from a location in Greece were identified as *C. grisebachii* ssp. *confusa* Hal cys, an endemic Greek species that is closely related, but was not treated in the recent revision, of the *C. stoebe* s.l. species group (Ochsmann, 2000). Samples from a location in Bulgaria were not able to be identified (*Centaurea* sp. unknown), but were considered as outside the *C. stoebe* species group.

Amplifying and sequencing

We explored several cpDNA introns using the universal primers of Taberlet *et al.* (1991) and Dumolin-Lapegue *et al.* (1997). We found the intergenic spacers between *trn*F (GAA) and the *trn*L
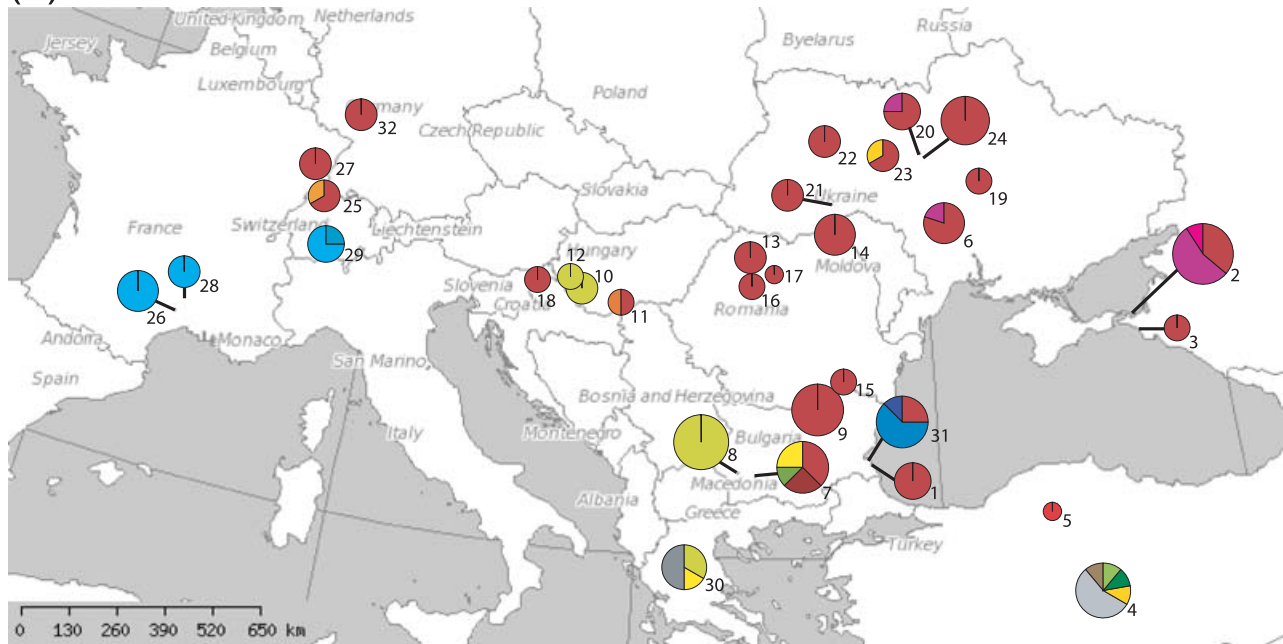
(a)
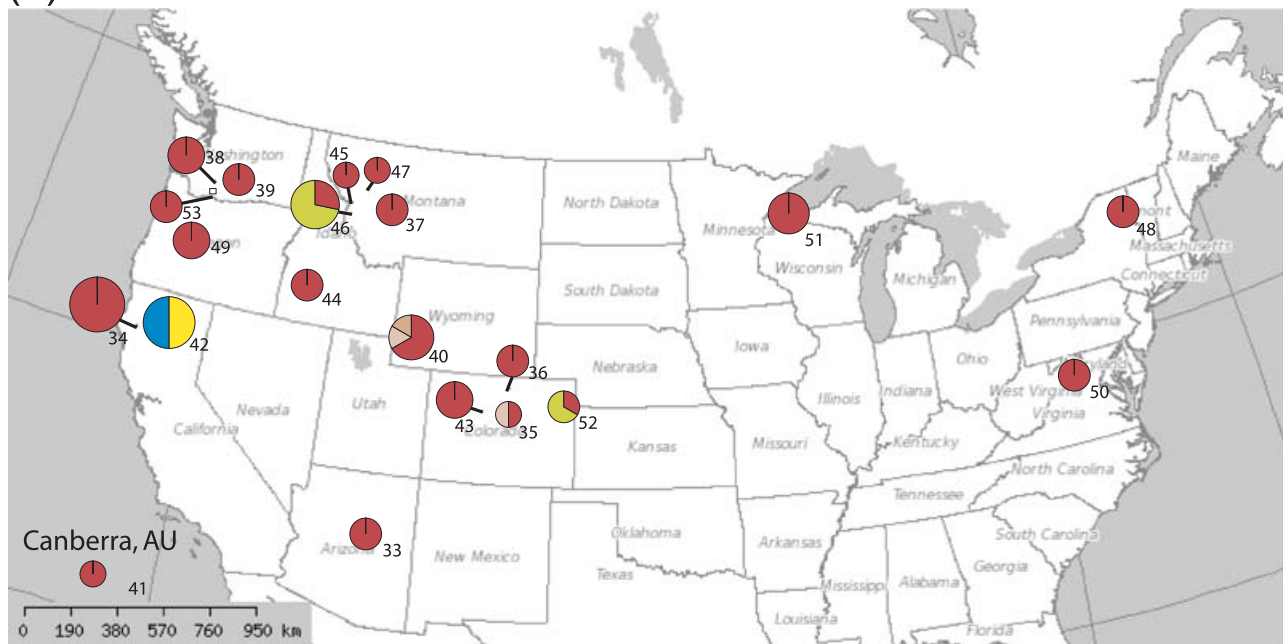
(b)

Figure 1 Map of sample sites. Pies showing haplotype frequency are sized proportionally to sample size. Site numbers correspond to those in Table 1, and colours correspond to the haplotype network in Fig. 2. (a) North America; (b) Europe.

(UAA) 3′ exon and between the *trn*L (UAA) 5′ exon and *trn*T (UGU) to be most variable (Hufbauer *et al*., 2004). We amplified the entire region from *trn*F to *trn*T in single reactions using the universal primer pair A50272 and B485557 of Taberlet *et al*. (1991). We sequenced these regions from one to 11 samples from each collection location (Table 1). The amplification program included an initial denaturation step at 94 °C for 1 min 30 s, then

30 cycles of 94 °C for 40 s, 53 °C for 40 s, and 72 °C for 2 min. After cycling, the program finished with a 10-min extension at 72 °C. Polymerase chain reaction (PCR) products were cleaned in an exonuclease/shrimp alkaline phosphatase incubation step (37 °C, 20 min, then 94 °C for 10 min to denature the enzymes) We sequenced from both ends of the PCR product into the variable intergenic spacers using BigDye Terminator Cycle

Table 1 Species, site locality names, site number, site codes, samples sizes, and haplotypes of individuals included in analyses.

Species	Site name	Site number	Site code	Sample size	Haplotypes
Samples from the native range					
Centaurea diffusa	Panitsovo, Bulgaria	1	BG34	4	H1
	Taiman Bay, Russia	2	RUS5	11	H1, H16, H17
	Anapa, Russia	33	RUS15	2	H1
	Ürgüp, Turkey	4	TR6	9	H7, H9, H11, H19, H22
	Çankiri, Turkey	5	TR10	1	H2
	Pervomaj'sk, Ukraine	6	UA3	5	H1, H16
C. stoebe micranthos	Jundola, Bulgaria	7	BG4	8	H1, H5, H8, H10
	Monastery Route, Bulgaria	8	BG8	9	H6
	Rüse, Bulgaria	9	BG27	8	H1
	Bohonye, Hungary	10	HU12	3	H6
	Batmonostor, Hungary	11	HU17	2	H1, H3
	Heviz, Hungary	12	HUH	2	H6
	Baia Mare, Romania	13	RO20	3	H1
	East of Ploiesti, Romania	14	RO24	5	H1
	Valea Argovei, Romania	15	RO25	2	H1
	Bicaz Chi, Romania	16	ROBC	2	H1
	Buru, Romania	17	ROB	1	H1
	Maribor, Slovenia	18	SLO6	2	H1
	Kholodne Jar, Ukraine	19	UA7	2	H1
	Bila Tserkva, Ukraine	20	UA15	4	H1, H16
	Kamjanec Podilsky, Ukraine	21	UA19	3	H1
	Ostrag, Ukraine	22	UA31	3	H1
	Berdiciv, Ukraine	23	UA35	3	H1, H11
	Kolotschke, Ukraine	24	UA 14/1/B	7	H1
C. stoebe stoebe	Basel, Switzerland	25	CH1	3	H1, H4
	Navacelles, France	26	F1	5	H13
	Kembs, France	27	F3	3	H1
C. maculosa albida	Anduze, France	28	F2	3	H13
C. vallesiaca	Valais, Switzerland	29	CH2	4	H12, H13
C. grisebachii ssp. *confusa*	Kastraki, Greece	30	GR50	6	H6, H10, H18
Centaurea ssp. *unknown*	Sozopol, Bulgaria	31	BG37	8	H1, H14, H15
C. xpsammogena	Mannheim, Germany	32	D1	3	H1
Samples from the introduced range*					
C. diffusa	Sedona, AZ	33	USAZ1	3	H1
	Trinity, CA	34	USCA2	9	H1
	Fort Collins, CO	35	USCO1	3	H1
	74E, CO	36	USCO2	2	H1, H20
	Helen, MT	37	USMT1	3	H1
	Bingen, WA	38	USWA1	4	H1
	Kittitas, WA	39	USWA2	3	H1
	Afton, WY	40	USWY1	6	H1, H20. H21
C. stoebe micranthos	Canberra, Australia	41	AUCA1	2	H1
	Log Jam, CA	42	USCA1	6	H10, H14
	Vail, CO	43	USCO3	4	H1
	Grimes Creek, ID	44	USID1	3	H1
	Florence, MT	45	USMT1	2	H1
	Hamilton, MT	46	USMT2	4	H1, H6
	Seely, MT	47	USMT3	2	H1
	Keene Valley, NY	48	USNY1	3	H1
	Bend, OR	49	USOR1	4	H1
	Middletown, VA	50	USVA1	3	H1
	Bayfield, WI	51	USWI1	5	H1
C. xpsammogena	Yuma, CO	52	USCO4	3	H1, H6
	Hood River, OR	53	USOR2	3	H1

*USA unless otherwise specified.

Primer code	Distance from B48557	Sequence 5′–3′
BF2	25	ATG CAT AGG AAT CAA TAA ACT CT
BF1	570	GGA TAT GGC GAA ATT GGT AGA C
BR2	715	CGC TTT CTG AAC CTT TGT TTG T
AR3	1420	ATG CCT CAT CCT CAT TTT C

Table 2 Primer codes, approximate distance (bp) relative to Taberlet *et al.* (1991) primer B48557, and 5′–3′ sequence for primers designed from *Centaurea* sequences for amplification and sequencing. All polymerase chain reactions annealed at 53 °C. In the primer codes, F refers to forward primers, and R to reverse.

Sequencing (version 3.1, Applied Biosystems, Foster City, CA, USA) primed with the PCR primers or internal primers designed from our *Centaurea* sequences (Table 2). Sequencing reactions were cleaned using ethanol precipitation, and nucleotides were separated on an ABI 3100 capillary instrument. If the entire sequenced region was not readable, or if there were single base-pair changes external to the haplotype network (see data analysis below), the PCR was repeated, and that individual was also sequenced from the opposite direction using an internal primer designed from *Centaurea* (Table 2). Sequences of each haplotype identified are available from GenBank (*trn*F (GAA)-*trn*L (UAA) 3′ exon: EF695029–EF695050; *trn*L (UAA) 5′ exon-*trn*T (UGU): EF695051–EF695072) Sequences were verified visually against the trace files in SeqMan (DNAStar, Inc., Madison, WI, USA); alignments were adjusted manually and trimmed to the same length. We joined sequences from the two segments together for the following analyses (independent analyses of the two segments, not shown, revealed similar patterns but with less resolution).

Data analysis

To evaluate levels of genetic variation in the native and introduced ranges we compared haplotype diversity (the probability that two samples' alleles are different) between native and introduced samples of *C. diffusa* and of *C. stoebe micranthos* using CONTRIB (Petit *et al.*, 1998). We based CONTRIB analyses on the smallest sample size of the groups being compared. To test the significance of these results, we constructed 95% confidence intervals from the standard error of estimated haplotype diversity. We also calculated allelic richness using CONTRIB. Allelic richness is an analogue of species richness and simply indicates the number of different alleles found in a sample. Like species richness, allelic richness is affected by sample size. CONTRIB accounts for different sample sizes using rarefaction (El Mousadik & Petit, 1996) to provide an intuitive number comparable across samples of difference size; however, CONTRIB does not provide statistics for estimates of the variance in allelic richness.

We constructed haplotype networks from sequence data using the network building software TCS 1.2.1 (Clement *et al.*, 2000), which uses statistical parsimony to construct an unrooted system of relationships between non-recombining sequences based on the genealogical reconstruction algorithm of Templeton *et al.* (1992). The nexus file is available in Appendix S1 in Supplementary Material. In these analyses, individual samples can be internal in the network. We treated indels as a fifth state (Giribet & Wheeler, 1999; Simmons & Ochoterena, 2000), and coded each indel as

a single mutational event (Simmons *et al.*, 2001). Initial analyses focused on *C. diffusa* and *C. stoebe micranthos* separately. We found that haplotypes were shared between the two, and also that some North American samples shared haplotypes with the other species (*Centaurea* sp. unknown and *C. grisebachii confusa*), and so we constructed a single haplotype network that included all taxa. We then performed a nested clade analysis to aid in distinguishing between gene flow, fragmentation, and range expansion (Templeton *et al.*, 1995, 1998). There has been some controversy over the effectiveness of nested clade analysis (Knowles & Maddison, 2002; Templeton, 2004), but it remains a common way to disentangle evolutionary processes leading to geographical patterns in population genetic data within and among closely related taxa. We employed the program ANECA (version 1.1, Panchal, 2007) for our nested clade analysis. ANECA implements both TCS and GEODIS (Posada *et al.*, 2000, 2006) and automates the inference process.

Finally, we performed maximum parsimony analyses in PAUP* version 4.0b10 (Swofford, 2003) using a branch and bound search. A majority-rule consensus tree was estimated, and relative reliability of groups was assessed using 20,000 replicates of fast-heuristic bootstrap resampling.

RESULTS

Alignments of the sequences showed several insertion/deletion sites (indels), and total sequence length ranged from 464 to 469 bp on the A50272 side (*trn*L), and from 385 to 404 on the B48557 side (*trn*T). In all, there were 26 variable sites in these regions among the sampled individuals, including nine indels and eight singletons.

Centaurea diffusa samples contained 11 haplotypes, nine of which were found in the native range, and three of which were found in the introduced range (Fig. 2), making haplotype diversity significantly higher in the native range (Table 3). Allelic richness of native *C. diffusa* was more than double that of introduced *C. diffusa* (Table 3).

Only one of the haplotypes harboured by *C. diffusa*, the most common (H1), was shared between individuals from the native and introduced range. This haplotype was found in about 68% of the *C. diffusa* samples: 44% of those from the native range, and 91% of those from the introduced range. It was found in all sample locations except for the two in Turkey. Two unique introduced haplotypes (H20, H21) were found in individuals from Wyoming (USWY1) and Colorado (USCO2), and were most similar to individuals from Ürgüp, Turkey (TR6, H19).
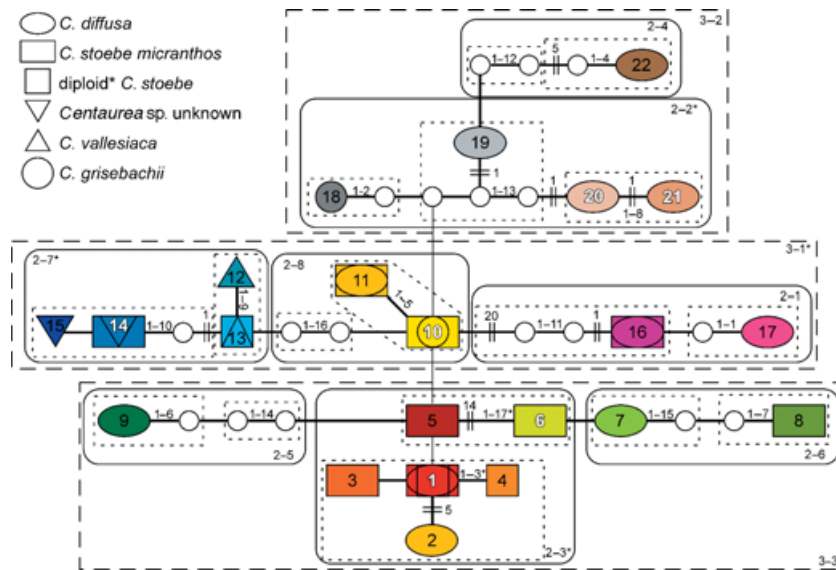
Figure 2 Parsimony network showing the relationships among cpDNA haplotypes from Table 1 and Fig. 1. Taxa are designated with individual shapes (see key), with diploid* *C. stoebe* indicating both *C. stoebe stoebe* and *C. maculosa albida*. Where more than one shape is present at a node in the network, more than one species harboured that haplotype. Small open circles indicate haplotypes that were not found in the samples. Locations of the nine indels are indicated with a double bar through the network, along with a number showing the length of the indel. Haplotype numbers in white indicate haplotypes found within the introduced range, all of which were also found in the native range except for H20 and H21. For the nested clade analysis, short-dashed boxes show one-step clades, with the clade number along a network branch where possible, solid boxes show two-step clades, and long-dashed boxes show three-step clades. Clades with significant geographical patters are noted with an asterisk. Inference key steps and conclusions are provided in Table 4.

Table 3 Haplotype diversity, number of haplotypes, and allelic richness of native and introduced (North American) samples of *Centaurea diffusa* and *Centaurea stoebe micranthos*.

Group	n	Mean haplotype diversity	Number of haplotypes	Allelic richness
Native *C. diffusa*	32	0.726	9	5.75
Introduced *C. diffusa*	33	0.170*	3	1.46
Native *C. stoebe micranthos*	69	0.500	8	4.55
Introduced *C. stoebe micranthos*	36	0.387	4	2.95

*Haplotype diversity of introduced populations of *C. diffusa* is significantly lower ($P < 0.01$) than native populations of the species.

Centaurea stoebe samples (including *C. stoebe stoebe*, *C. stoebe micranthos*, and *C. maculosa albida* within the native range) harboured 11 haplotypes, 10 of which were found in the native range, and four of which were found in the introduced range (Figs 1 & 2). Focusing only on *C. stoebe micranthos*, eight haplotypes were found in the native range and four in the introduced range. While haplotype diversity is higher in the native range than the introduced range for *C. stoebe*, the difference is not significant (Table 3). Allelic richness of native *C. stoebe* was 1.7 times that of the samples from the introduced range (Table 3).

Three of the four haplotypes found within the introduced samples of *C. stoebe micranthos* also were present within the samples of *C. stoebe micranthos* from the native range. The fourth haplotype found in North American samples of *C. stoebe micranthos* (H14) was found in the native range in the samples

of *Centaurea* sp. unknown from Sozopol, Bulgaria. Those haplotypes that were found in *C. stoebe micranthos* in both ranges included the two most common (H1, H6), and one other (H10). The most common haplotype (H1) was found in 62% of the *C. stoebe micranthos*: 52% of those from the native range, and 79% of those from the introduced range. All but three sample locations of *C. stoebe micranthos* in the native range and one from the introduced range contained H1.

The second most common haplotype in the network (H6) was shared between 14 native and four introduced individuals. This haplotype was found in the native range in two *C. stoebe micranthos* sample locations (from Bulgaria and Hungary), and in *C. grisebachii confusa* (Greece). In the introduced range it was found in *C. stoebe micranthos* from Montana, and in *C. xpsammogena* from Colorado. The third most common haplotype in the network (H14) was shared between four individuals from

Table 4 Clades, steps followed in the inference key, and conclusion from the nested clade analysis for clades with significant geographical patterns (see Fig. 2). Nested clades within each higher step clade with significant distance values are indicated. D_C is a measure of the geographical spread of a clade. D_N measures how a clade is geographically distributed relative to other clades at the level of nesting. The superscripts S and L indicate whether the distance is significantly small or large, respectively, at the $P < 0.05$ level. For details of statistics see Appendix S2 in Supplementary Material.

Clade	Inference key steps	Conclusion	Nested clades with significant distances
2–2	1-19-20-2-11-12 No	Contiguous range expansion	1–2: D_N^S
			1–8: D_N^L
			1–13: D_C^S
2–3	1-2-11-12 No	Contiguous range expansion	1–17: D_C^S, D_N^S
2–7	1-19 No	Allopatric fragmentation	1–9: D_C^S, D_N^S
			1–10: D_C^L, D_N^L
3–1	1-2-3-4 No	Restricted gene flow with isolation by distance	2–1: D_C^S
3–3	1-2-3-4 No	Restricted gene flow with isolation by distance	2–6: D_C^S
Total	1-2-11-12 No	Contiguous range expansion	3–1: D_C^S, D_N^S

California (USCA1), and five individuals of *Centaurea* sp. unknown from Sozopol, Bulgaria (BG37). Thus, in combination, the Bulgarian samples harbour every haplotype found in the North American samples.

The haplotype network in Fig. 2 shows that three haplotypes are shared between *C. diffusa* and *C. stoebe micranthos* (H1, H11, and H16). The individuals thought to be of hybrid origin contained the most common haplotype (H1) or the second most common haplotype (H6), that otherwise was associated only with *C. stoebe micranthos*.

The nested clade analysis (Fig. 2, Table 4, details are available in Appendix S2 in Supplementary Material) shows significant geographical patterns. The signature of the invasions shows up as inferences of allopatric fragmentation or range expansion for two-step clades including samples from both the native and the introduced ranges (e.g. 2–2, 2–3, 2–7). Additionally, the entire network showed a pattern of contiguous range expansion (Table 4).

The maximum parsimony tree (Fig. 3) supports the same basic patterns found in the haplotype network, with strong support for several of the clades identified by the nested clade analysis.

DISCUSSION

The introns between *trn*L and *trn*T provide several important pieces of information key to understanding the invasions of *C. diffusa* and *C. stoebe micranthos* into North America. For *C. diffusa*, the data suggest that at least two introductions have occurred: one of the most common haplotype (H1) from one or more of the many locations within the native range that contain that haplotype, and another from a location containing the related haplotypes H20 and H21. The closest native haplotype to those two introduced haplotypes is common in Ürgüp, Turkey (H19). It is possible, however, that there exists a region in the native range from which we did not sample that contains both the most common haplotype and the unique North American haplotypes. A single introduction from such an area could account for our findings.

Despite the likelihood of at least two introductions *C. diffusa* appears to have lost variation in cpDNA haplotypes during its invasion of North America. This loss could be due to founder effects, selection upon arrival, or drift due to initially small populations.

The data suggest that *C. stoebe micranthos* also was introduced more than once. It is conceivable, however, that those introductions were from a relatively narrow geographical region, given that samples from Bulgaria contained all of the haplotypes we found in North American *C. stoebe micranthos*. Interestingly, some of those Bulgarian samples were not *C. stoebe micranthos*, but rather an unknown species outside the larger *C. stoebe* species group. No single sample location within our Bulgarian samples contained all the North American haplotypes. The diversity we found in the introduced range is therefore most easily explained by multiple introductions from that region. Indeed, that area of central Europe was an important source of alfalfa seed shipped into North America (Roché & Roché, 1991), which makes multiple introductions seem feasible. While introduced individuals match Bulgarian samples well, introductions may also have occurred from other areas. For example, the common haplotype (H1) was found in most of the native locations sampled and it could be from one or more of them.

Haplotype diversity of *C. stoebe micranthos* was not significantly lower in the introduced range relative to the native range. The introduced North American *C. stoebe micranthos* samples harbour a haplotype that we did not find within our *C. stoebe micranthos* samples from the native range, but did find within a different taxon. The presence of this haplotype elevates haplotype diversity in North America.

The finding that California harbours a haplotype (H14) that matches a different taxon (*Centaurea* sp. unknown) suggests the possibility of a cryptic invasion particularly given the diversity and complexity of *Centaurea* in Eurasia. Sharing of haplotypes was found among other taxa as well, however. For example, *C. stoebe micranthos* shared haplotypes with *C. diffusa*, *C. stoebe stoebe*, and *C. grisebachii*. The frequency with which the same sequence was found in different taxa suggests incomplete
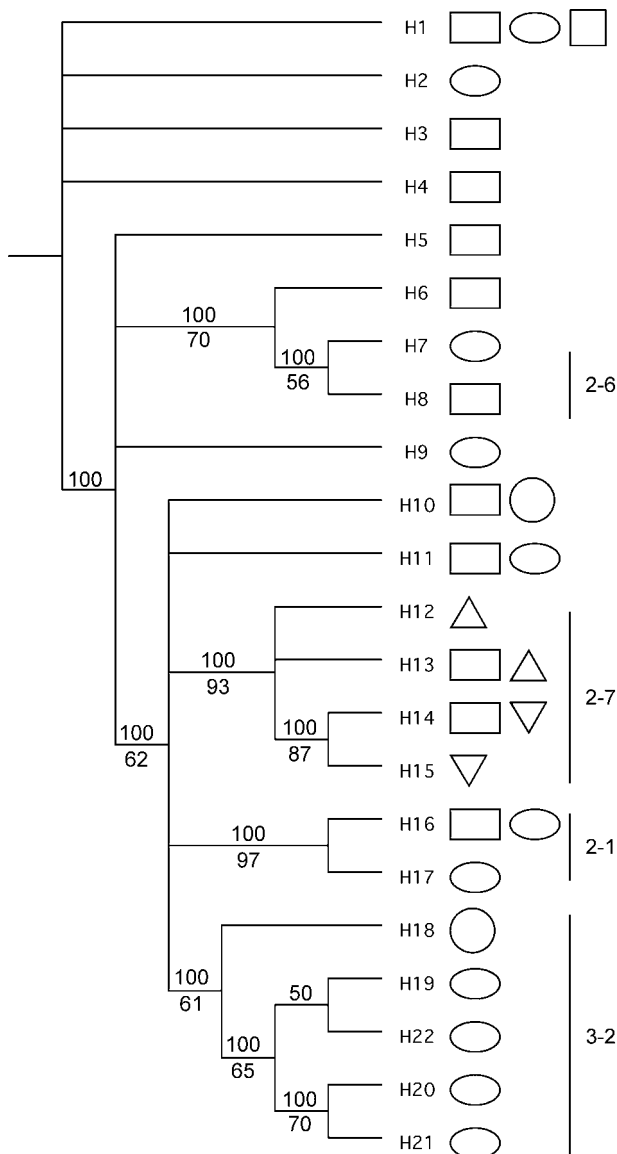
Figure 3 Majority-rule consensus tree based on the two most-parsimonious trees found in a maximum parsimony branch and bound search of the 22 haplotypes. Numbers above branches indicate majority-rule percentages, those below the branch indicate bootstrap values > 50%. Taxa are indicated by shape following the legend in Fig. 2. Clades in common with the nested clade analysis from Fig. 2 are also noted.

Clearly, as with others cpDNA intergenic spacers and ITS (Garcia-Jacas *et al.*, 2006), the *trn*F-trnT intergenic spacers do not provide enough resolution to separate closely related *Centaurea* taxa clearly. Rather, these sequence data seem to provide more geographical information than taxon-level information. More variable markers, such as microsatellites and amplified fragment length polymorphisms may distinguish both taxa and geographically distinct populations more effectively.

Despite limited variation, the data presented here help narrow down the sources of the invasions into North America of *C. diffusa* and *C. stoebe micranthos*. These findings can direct evaluations of biogeographical hypotheses to explain invasions (e.g. enemy release and EICA) mainly through excluding regions that are unlikely to serve as the origin of the invasions. For example, comparing North American plants to those from Navacelles, France, which did not harbour the most common haplotype and is therefore not likely to represent a location that contributed to the founding of the invasion, would not be appropriate. In contrast, comparisons between locations in Bulgaria and North America would be more suitable. Finally, the data from North American samples can be used to ensure that adequate genotypic sampling is used in testing of host range and efficacy, if additional biological control introductions are considered.

ACKNOWLEDGEMENTS

REFERENCES

Blank, R.R. & Sforza, R. (in press) Plant-soil relationships of the invasive annual grass *Taeniatherum caput-medusae*: a reciprocal transplant experiment. *Plant and Soil*, **298**, 7–19.

Blossey, B. & Nötzold, R. (1995) Evolution of increased competitive ability in invasive nonindigenous plants – a hypothesis. *Journal of Ecology*, **83**, 887–889.

Callaway, R.M., Thelen, G.C., Barth, S., Ramsey, P.W. & Gannon, J.E. (2004) Soil fungi alter interactions between the invader *Centaurea maculosa* and North American natives. *Ecology*, **85**, 1062–1071.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1687–1699.

Dumolin-Lapegue, S., Pemonge, M.-H. & Petit, R.J. (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393–397.

segregation of haplotypes that were present in the common ancestor of these species, or hybridization moving cpDNA haplotypes between taxa. Probably both mechanisms are acting, as hybridization is likely to be common in this genus between taxa with compatible ploidy levels, but rare between taxa like *C. stoebe micranthos* and *C. diffusa*, with different ploidy levels. In this slowly evolving intergenic spacer, homoplasy does not seem a likely explanation for the sharing of these haplotypes between taxa, particularly given that the clade was found in preliminary work exploring other markers, as well (Hufbauer *et al.*, 2004).

El Mousadik, A. & Petit, R.J. (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.

Ellstrand, N.C. & Schierenbeck, K.A. (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences USA*, **97**, 7043–7050.

Facon, B., Genton, B.J., Shykoff, J.A., Jarne P., Estoup A. & David, P. (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution*, **21**, 130–135.

Garcia-Jacas, N., Uysal, T., Romashchenko, K., Suarez-Santiago, V.N., Ertugrul, K. & Susanna, A. (2006) *Centaurea* revisited: a molecular survey of the Jacea group. *Annals of Botany*, **98**, 741–753.

Gaskin, J.F., Zhang, D.Y. & Bon, M.C. (2005) Invasion of *Lepidium draba* (Brassicaceae) in the western United States: distributions and origins of chloroplast DNA haplotypes. *Molecular Ecology*, **14**, 2331–2341.

Gáyer, G. (1909) Vier neue Centaureen der Flora von Ungarn. *Magyar Botanikai Lapok*, **8**, 58–61.

Giribet, G. & Wheeler, W.C. (1999) On gaps. *Molecular Phylogenetics and Evolution*, **13**, 132–143.

Goolsby, J.A., De Barro, P.J., Makinson, J.R., Pemberton, R.W., Hartley, D.M. & Frohlich, D.R. (2006) Matching the origin of an invasive weed for selection of a herbivore haplotype for a biological control programme. *Molecular Ecology*, **15**, 287–297.

Hufbauer, R.A., Marrs, R.A., Jackson, A.K., Sforza, R., Bais, H.P., Vivanco, J.M. & Carney, S.E. (2004) Population structure, ploidy levels and allelopathy of spotted and diffuse knapweed in North America and Eurasia. *Proceedings of the XI international symposium on biological control of weeds* (ed. by J.M. Cullen, D.T. Briese, D.J. Kriticos, W.M. Lonsdale, L. Morin and J.K. Scott), pp. 121–126. CSIRO Entomology, Canberra, Australia.

Jahodová, S., Trybush, S., Pyšek, P., Wade, M. & Karp, A. (2007) Invasive species of *Heracleum* in Europe: an insight into genetic relationships and invasion history. *Diversity and Distributions*, **13**, 99–114.

Knowles, L.L. & Maddison, W.P. (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Kolbe, J.J., Glor, R.E., Schettino, L.R.G., Lara, A.C., Larson, A. & Losos, J.B. (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Lavergne, S. & Molofsky, J. (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences USA*, **104**, 3883–3888.

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. & Bazzaz, F.A. (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.

Maddox, D.M. (1979) The knapweeds: their economics and biological control in the western states, USA. *Rangelands*, **1**, 139–141.

Marrs, R.A., Hufbauer, R.A., Bogdanowicz, S.M. & Sforza, R. (2006) Nine polymorphic microsatellite markers in *Centaurea stoebe* L. [subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek] and *C. diffusa* Lam. (Asteraceae). *Molecular Ecology Notes*, **6**, 897–899.

Mitchell, C.E., Agrawal, A.A., Bever, J.D., Gilbert, G.S., Hufbauer, R.A., Klironomos, J.N., Maron, J.L., Morris, W.F., Parker, I.M., Power, A.G., Seabloom, E.W., Torchin, M.E. & Vazquez, D.P. (2006) Biotic interactions and plant invasions. *Ecology Letters*, **9**, 726–740.

Müller, H. (1989) Growth pattern of diploid and tetraploid spotted knapweed, *Centaurea maculosa* Lam (Compositae), and effects of the root-mining moth *Agapeta zoegana* (L.) (Lep, Cochylidae). *Weed Research*, **29**, 103–111.

Ochsmann, J. (2000) Morphologische und molekularsystematische Untersuchungen an der *Centaurea stoebe* L.-Gruppe (Asteraceae-Cardueae) in Europa. *Dissertationes Botanicae*, **324**, 1–33.

Panchal, M. (2007) The automation of nested clade phylogeographic analysis. *Bioinformatics*, **23**, 509–510.

Parker, I.M., Simberloff, D., Lonsdale, W.M., Goodell, K. & Wonham, M. (1999) Impact: toward a framework for understanding the ecological effects of invaders. *Biological Invasions*, **1**, 3–19.

Pearson, D.E., McKelvey, K.S. & Ruggiero, L.F. (2000) Non-target effects of an introduced biological control agent on deer mouse ecology. *Oecologia*, **122**, 121–128.

Petit, R.J., El Mousadik, A. & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000) Environmental and economic costs of nonindigenous species in the United States. *Bioscience*, **50**, 53–65.

Posada, D., Crandall, K.A. & Templeton, A.R. (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Posada, D., Crandall, K.A. & Templeton, A.R. (2006) Nested clade analysis statistics. *Molecular Ecology Notes*, **6**, 590–593.

Pyšek, P. & Richardson, D.M. (2006) The biogeography of naturalization in alien plants. *Journal of Biogeography*, **33**, 2040–2050.

Richardson, D.M. & Pyšek, P. (2006) Plant invasions – merging the concepts of species invasiveness and community invasibility. *Progress in Physical Geography*, **30**, 409–431.

Roché, B.F. & Roché, C.T. (1991) Identification, introduction, distribution, and economics of *Centaurea* species. *Noxious range weeds* (ed. by L.F. James, J.O. Evans, M.H. Ralphs and R.D. Child), pp. 274–291. Westview Press, San Francisco, California.

Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., Baughman, S., Cabin, R.J., Cohen, J.E., Ellstrand, N.C., McCauley, D.E., O'Neil, P., Parker, I.M., Thompson, J.N. & Weller, S.G. (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305–332.

Seastedt, T.R., Gregory, N. & Buckner, D. (2003) Effect of biocontrol insects on diffuse knapweed (*Centaurea diffusa*) in a Colorado grassland. *Weed Science*, **51**, 237–245.

Shea, K. & Chesson, P. (2002) Community ecology theory as a framework for biological invasions. *Trends in Ecology & Evolution*, **17**, 170–176.

Sheley, R.L., Jacobs, J.S. & Carpinelli, M.L. (1999) Spotted knapweed. *Biology and management of noxious rangeland weeds* (ed. by R.L. Sheley and J.K. Petroff), pp. 350–361. Oregon State University Press, Corvalis, Oregon.

Simmons, M.P. & Ochoterena, H. (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, **49**, 369–381.

Simmons, M.P., Ochoterena, H. & Carr, T.G. (2001) Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Systematic Biology*, **50**, 454–462.

Story, J.M., Good, W.R., White, L.J. & Smith, L. (2000) Effects of the interaction of the biocontrol agent *Agapeta zoegana* L. (Lepidoptera: Cochylidae) and grass competition on spotted knapweed. *Biological Control*, **17**, 182–190.

Suding, K.N., LeJeune, K.D. & Seastedt, T.R. (2004) Competitive impacts and responses of an invasive weed: dependencies on nitrogen and phosphorus availability. *Oecologia*, **141**, 526–535.

Swofford, D.L. (2003) PAUP* *phylogenetic analysis using parsimony (*and other methods)*, Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Templeton, A.R. (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton, A.R. (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789–809.

Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic–analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. 3. Cladogram estimation. *Genetics*, **132**, 619–633.

Templeton, A.R., Routman, E. & Phillips, C.A. (1995) Separating population structure from population history: a cladistic-analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

Thuiller, W., Richardson, D.M., Pyšek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.

Torchin, M.E. & Mitchell, C.E. (2004) Parasites, pathogens, and invasions by plants and animals. *Frontiers in Ecology and the Environment*, **2**, 183–190.

Watson, A.K. & Renney, A.J. (1974) The biology of Canadian weeds: *Centaurea diffusa* and *C. maculosa*. *Canadian Journal of Plant Science*, **54**, 687–701.

Editor: David Richardson

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Table S1 provides a nexus file of the 22 haplotypes.

Appendix S2 Table S2 provides detailed nested clade analysis statistics.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1472-4642.2007.00424.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.